south dakota
STATE HISTORICAL SOCIETY
DEPARTMENT OF EDUCATION

Digitization/Scanning Standards for State Agencies

If your agency's retention manual instructs certain documents to be kept digitally, you are responsible for making sure that those files are able to be opened for the entire life of the document, which in some cases may mean your files must be able to be opened 100 years from now. Consider how technology changes and the burden becomes apparent. Starting with a good quality scan can make this job easier.

Standard Textual Documents

Includes documents created using a word processing tool, such as typewriter or computer

- Save file as PDF/A
- Scan in bitonal black and white, not in grayscale
- Set resolution to 300 pixels per inch (ppi) to allow for sufficient clarity and to support the use of optical character recognition (OCR)
- Run OCR processing if desired

Low Quality Textual Documents

Includes documents with low contrast between text and the paper, that are handwritten, or are otherwise illegible

- Save file as PDF/A
- Scan in 8-bit grayscale or 24-bit RGB or CYMK color mode, depending on which best captures details and information
- Set resolution to 300 ppi minimum, up to 600 ppi for additional detail

Image Documents

Includes documents that are graphic in nature, such as maps, photographs, blueprints, engineering plans, charts, etc.

- Save images as TIFF
- If compression is required to reduce file size, lossless rather than lossy compression must be used
- Scan black and white photographic prints in 8-bit grayscale; if the image is faded or damaged, scan in 16-bit grayscale or 24-bit color
- Scan color images of any kind in 24-bit RGB or CMYK color
- The minimum resolution that any image should be scanned at is 300 ppi, but 400-600 ppi is recommended.
- For smaller photographs or negatives, or for images where more detail needs captured, scan at a higher resolution (example: scan 35mm negatives at 2800 ppi)

In-depth looks at….

Resolution: the number of pixels per inch (ppi) created during document capture. The higher the resolution, the more detail is captured. On some scanning hardware, this may be shown as dpi, or dots per inch, especially where the hardware also has a printing function.

PDF/A: an ISO-standardized version of the .pdf file type that is specialized for archiving and long term preservation.

Optical Character Recognition (OCR): an optional post-scan process that makes the document text-searchable.

Pixel bit depth: the number of color shades that can be represented by the amount of information saved for each pixel. A pure black and white image depth is 1-bit per pixel, while color images should be set to 24-bit. For color modes, RGB is suitable for digital viewing only, but if files will ever be printed, use CYMK.

Compression: an encoding process that changes how data is stored to reduce file size. Lossy compression discards data; lossless compression does not. LZW and JPEG2000 are acceptable lossless compression algorithms.

Open format: a file format not protected by intellectual property rights; more likely to retain openability further into the future.

Have questions? Want more information?

South Dakota State Archives: 773-4369

nicole.hosette@state.sd.us

Records Management: 773-3589

boa.records.frontdesk@state.sd.us

Thanks to the New York State Archives: http://www.archives.nysed.gov/common/archives/files/mr_erecords_imgguides.pdf

south dakota STATE HISTORICAL SOCIETY
DEPARTMENT OF EDUCATION

Digitization/Scanning Standards for State Agencies

Quick Reference Chart

Type	**Description**	**File Format** *Dictates how you save the finished scan*	**Resolution (in pixels per inch)**[1] *Dictates amount of detail in scan*	**Bit depth**[3] *Dictates number of color shades saved per pixel*
Standard text	Documents created with a word processing tool, like MS Word or a typewriter	PDF/A-1 or PDF/A-2 (.pdf)	72-300 ppi (depending on complexity of doc, how long it is to be	Bitonal black and white
Hybrid text	Text with figures, charts, handwriting	PDF/A-1 or PDF/A-2 (.pdf)	300 ppi	Bitonal b&w, 8-bit grayscale, or 24-bit color
Low quality text	Low contrast between text and paper; handwritten; triplicates; illegible	PDF/A (.pdf) or TIFF (.tif)[2]	Up to 600 ppi for additional detail	8-bit grayscale or 24-bit color
Black and white image	Photograph prints,	TIFF (.tif)	400-600 ppi	8-bit grayscale
Color image	Photograph prints, artwork, detailed charts	TIFF (.tif)	400-600 ppi	24-bit color
Maps	Full color geographic/spatial documents	TIFF (.tif) or PDF/A (.pdf)	300-600 ppi	24-bit color
Blueprints/plans	Includes engineering plans, diagrams, etc. Usually up to 2 colors	TIFF (.tif) or PDF/A (.pdf)	300-600 ppi	Bitonal b&w, 8-bit grayscale, or 24-bit color
Negatives/slides	35 mm film negatives; 120 or 220 slides	TIFF (.tif)	1600-2800 ppi (the smaller the image, the larger the ppi is needed)	8-bit grayscale or 24-bit color; reverse polarity for negatives

[1] Also represented as dots per square inch, or dpi. The same numeric value listed above applies to both.

[2] Images are recommended to be saved uncompressed. If file size is an issue, use a lossless compression standard, such as JPEG 2000 or LZW.

[3] Some hardware automatically detects color depth. Where multiple standards are listed, choose which one most closely aligns to the document content being digitized.